

Katja Siekmann

**Grund- und Orientierungswortschatz
für die Primarstufe**

**Häufigkeitsbasierter Wortschatz,
Phonem-(Basis-)Graphem-Korrespondenzen,
Fehlerverteilungen und didaktische
Implikationen**

**Band 1
Konsonanten**

Siekmann Verlag

Für meinen Bruder Dirk

&

Bereits vorausgegangen, aber nie vergessen: Heinz Siekmann. Danke für alles!

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

Cover & Illustrationen: Laura Dajana, Frankfurt a. M. (www.LauraDajana.com)

© 2023 Siekmann Verlag, Schönau am Königssee
Prof. Dr. Katja Siekmann (www.katja-siekmann.de)
2. überarb. Aufl.

Das Werk und seine Teile sind urheberrechtlich geschützt. Jede Verwendung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung der Autorin unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen oder die Einspeicherung und Verarbeitung in elektronischen Systemen. Dies gilt auch für Intranets von Schulen und sonstigen Bildungseinrichtungen.

Herstellung: BoD – Books on Demand, Norderstedt

ISBN: 978-3-9823973-9-9

Band I: Konsonanten

Einleitung _____ 5

Wortschatz, Phonem-Graphem-Relation(en) und Fehlerverteilung
zu:

(1) /k/+v/ _____ 15

1 /ks/ _____ 20

2 /pf/ _____ 26

3 /j/ _____ 31

4 /ŋ/ _____ 37

5 /p/ _____ 44

6 /ts/ _____ 52

7 /h/ _____ 61

8 /f/ _____ 66

9 /b/ _____ 74

10 /g/ _____ 80

11 /k/ _____ 85

12 /v/ _____ 97

13 /z/ _____ 102

14 /f/ _____ 106

15 /ç/ und /x/ _____ 116

16 /m/ _____ 122

17 /l/ _____ 128

18 /s/ _____ 135

19 /d/ _____ 144

20 /r/ _____ 149

21 /t/ _____ 156

22 /n/ _____ 167

Literaturverzeichnis _____ 175

Überblick: Phonem-Graphem-Verhältnisse *Konsonanten* _____ 179

Überblick: Ph-G-Verhältnisse *Konsonanten & Vokale* _____ 181

Liste der 100 häufigsten Wörter und Zusammenschluss _____ 184

Facciamo sia più semplice!
(Machen wir es einfacher!)

Danksagung

An dieser Stelle möchte ich Prof. Dr. Günther Thomé und Dr. Dorothea Thomé für die jahrelange Unterstützung und Bestärkung danken. Diese Auszählung stellt eine Erweiterung der (erstmalig vor einem Jahrzehnt veröffentlichten) gemeinsamen Häufigkeitsauszählung dar und soll die Ergebnisse noch greifbarer für die schulische und außerschulische Praxis aufbereiten.

Mein Dank gilt zudem den ehemaligen (Projekt-)Mitarbeiterinnen, Laura Dajana (Illustrationen & Cover) und Daniel Autenrieth (Homepage) für die sehr gelungenen Arbeiten und allen, die mit mir in moralischer, kollegialer und/oder freundschaftlicher Weise verbunden sind und an mich und dieses Projekt glauben.

Diese 2. überarb. Auflage enthält wenige Änderungen/Ergänzungen durch die Arbeit an Band 2.

Katja Siekmann
Schönau am Königssee im November 2023

Einleitung

Die statistische Auszählung von Wortschätzen als auch die Auszählung von Fehlschreibungen¹ haben eine über hundertjährige Tradition. Auf der Basis dieser Ergebnisse sind u.a. standardisierte und informelle Rechtschreibdiagnoseverfahren entstanden, aber auch die neueren Grund-/Mindestwortschätze der Bundesländer in Deutschland, ungeachtet dessen, dass keine *aktuelle* Häufigkeitsauszählung zu dem genutzten Schriftwortschatz von Schülerinnen und Schülern (im Folgenden: SuS²) vorliegt (vgl. u.a. Siekmann 2021; Hoffmann-Erz 2019).

Die vorliegende Arbeit soll diese Lücke auf spezielle Weise schließen: Es geht nicht nur um die Auszählung des Wortschatzes und der enthaltenen Fehlschreibungen. Es geht vielmehr darum, den Wortschatz auf der kleinsten sprachlichen Ebene zu betrachten und davon ausgehend didaktische Grundüberlegungen für eine konzeptionelle Neugestaltung des Anfangsunterrichts anzubieten.

Warum sollten Häufigkeiten im Schreibwortschatz von SuS im (Anfangs-)Unterricht beachtet werden?

Die deutsche Schriftsprache ist regulär aufgebaut und sollte SuS nach einem aufbauenden Prinzip vermittelt werden (Thomé & Thomé 2023). Aufbauend im Sinne von Comenius (1657/1992)

¹ Wortschatz: Mündliche oder schriftliche Korpora, meist Erwachsenensprache, siehe Übersicht u.a. in Naumann (1999); Fehlerforschung: Übersicht in Siekmann & Thomé (2018).

² Müsste es vor dem Hintergrund der Auszählung durch „SchuSch“ abgekürzt werden? Näheres dazu im Kapitel zum Phonem /f/ (S. 66f.).

bedeutet, dass vom „**Einfachen zum Komplexen/ Schwierigen**“ gelernt werden sollte.

Was kann bei der Schriftsprache als „einfach“ und was als „schwierig“ bezeichnet werden?

SuS gehen zunächst von der Lautung eines zu schreibenden Wortes aus und müssen entsprechende Repräsentanten auf der Schriftebene kennenlernen. Eine größere sprachstatistische Analyse zu den Phonem-Graphem-Relationen zeigte, dass es grundlegende und **häufige Repräsentationen für ein Phonem** gibt (**Basisgrapheme, kurz: BG**) und mintunter auch andere, **statistisch seltenere Grapheme**, die sich auf dasselbe Phonem beziehen, die sog. **Orthographeme** (kurz: **OG**).³

90 % aller Grapheme in Texten sind Basisgrapheme.

Seltene Schreibungen wie die Doppelkonsonanz oder die Längemarkierung der Vokale bilden zusammen rund 5 %, die Wahrung des morphematischen Schemas (Wortstammschreibungen) bzw. historische Formen bilden ebenfalls lediglich 5 % der Grapheme ab (**Orthographeme** somit insgesamt **10 %**; vgl. u.a. auch Thomé 2019, Corvacho del Toro 2017, Corvacho del Toro & Thomé 2021).

Die **deutsche Schriftsprache** weist zudem eine sog. ‚**degressive Struktur**‘ auf. Degressiv bedeutet, dass verhältnismäßig wenige Wörter einen großen Teil deutscher Normaltexte ausmachen. Nach Spitta (2000) decken die 100 häufigsten Wörter ca. 50 %,

³ vgl. Thomé, Siekmann & Thomé (2011/2016; 100.000er-Auszählung, Erwachsenentexte). Thomé hierzu bereits in den 90er Jahren; ebenso Naumann, der für die Unterscheidung die Bezeichnungen *Mehrheits-* und *Minderheitsschreibungen* prägte (vgl. auch Naumann 2017).

die 1.000 häufigsten Wörter 80 % eines Textes ab. Die meisten der 100 häufigsten Wörter [Spitta führt die Meier'sche Zählung von 1967 (!) an, der sich auf Kaeding (1898) bezog] sind **Struktur- oder Funktionswörter** wie *Artikel, Pronomen* und *Konjunktionen* (vgl. u.a. Augst & Dehn 2009, Siekmann 2022a,b & 2023a,b). Für die Rechtschreibung bedeutet dies: **Wird die Orthographie dieser Wörter beherrscht, ist ein erheblicher Teil eines Textes korrekt verschriftet.** In dieser Auszählung liegt der prozentuale Anteil der **Top 100** sogar bei **rund 56 %** (vgl. S. 184; auch in Siekmann 2023a,b).

Aus didaktischer Sicht wäre eine Mischung aus häufigen Wörtern (Struktur- und Funktionswörter) sowie (über-)individuellen Lernwörtern mit orthographisch aufbauendem Charakter (von Basis- zu Orthographemen/häufige Modell- und Ankerwörter) sinnvoll. Anzustreben wäre ein Anfangsunterricht, der die Häufigkeiten im Schreibwortschatz beachtet und einer entwicklungsorientierten und lernpsychologisch logischen Progression folgt. Materialien für den Anfangsunterricht bieten diese basale Orientierung bisher nicht, vermutlich auch, weil die Sprachdidaktik der empirischen Grundlagenforschung bei der Entwicklung von Vermittlungsansätzen bisher wenig Beachtung geschenkt hat.

Forschungsfragen

- Die grundlegenden Forschungsfragen dieser Auszählung lauten:
- Welche Phonem-Graphem-Relationen sowie Wörter/Wortbausteine sind häufig und sollten im Anfangsunterricht (auch im Hinblick auf die Textproduktion) priorisiert werden?
 - Sind die seltenen und weniger häufigen (Ortho-)Grapheme im Vergleich zu den häufigsten Graphemen fehleranfälliger?

Welche sollten mit Blick auf die Textproduktion frühzeitig behandelt werden?

- Finden sich Rechtschreibphänomene wie u.a. die Länge-/Kürzemarkierung in einer großen Streubreite an Wörtern und/oder lassen sich überindividuelle Modell-/Ankerwörter eruieren?

Aus den gewonnenen Erkenntnissen sollen didaktische Implikationen für die Förderung in der Praxis abgeleitet werden. Es handelt sich somit um eine praxisorientierte, sprachwissenschaftliche Grundlagenforschung, die der Professionalisierung von schulischen und außerschulischen Lehrkräften dienen soll (vgl. auch Siekmann 2021). Auch Schulbuchverlage sind eingeladen, ihre Materialien in Bezug auf die Ergebnisse zu überdenken.

Korpus

Das Korpus besteht aus rund 1.000 Texten von knapp 250 SuS der 3.-5. Jahrgangsstufe (neun komplette Klassen unterschiedlicher Schulen aus NRW). Bei den Schreibprodukten handelt es sich überwiegend um freie Texte (Erlebniserzählungen, Briefe, Phantasiegeschichten, Berichte usw.), einige wenige Texte entstanden auf der Grundlage von sehr offenen Bildimpulsen.⁴ Es gab keine orthographischen, lexikalischen oder textstrukturellen Hilfestellungen.

Die Texte wurden digitalisiert (Abschrift mit und ohne Fehlschreibungen) und im Anschluss manuell nach der Zielsetzung der Analyse quantitativ und qualitativ ausgezählt (Wortschatz

⁴ Was bedeutet „frei“ im Rahmen einer (schulischen) Institution? – Die Frage wird in Diskussionen um die Auszählung (zu Recht) gestellt. Freie Textprodukte von 250 SuS außerhalb der Institution zu erfragen, wäre jedoch organisatorisch schwer umsetzbar gewesen und die Lehrkräfte erhielten zumindest detaillierte Lernstandsrückmeldungen.

allgemein/Wortschatz zusammengefasst, sprich: Flexionsformen zusammengezogen; Phonem-Graphem-Korrespondenz aller Relationen inkl. der jeweiligen Fehlschreibungen). Namen wurden während der Digitalisierung bereits gestrichen (analog zur 2011er-Auszählung), da sie nicht regulär nach den Prinzipien der deutschen Schrift gebildet werden (müssen). Eine automatisierte Auszählung war aufgrund von möglichen Fehlinterpretationen zwischen Buchstaben-/Graphem-Ebene nur bedingt möglich.

Beispiel: Der Buchstabe „h“ ist Basisgraphem <h> für /h/ (Haus). Automatisiert würde das Schriftzeichen aber auch in Längemarkierungen wie <ah> (Zahn) erfasst, wo es keinen eigenen Lautwert besitzt.

In der Wortschatzauszählung verteilen sich insgesamt rund 122.000 Wörter (141.000 inkl. Namen) auf 6.741 unterschiedliche Worteinträge. Um die Phonem-Graphem-Ebene zu betrachten sind sämtliche Flexionsformen einzeln erfasst.

Auswertung

Die Häufigkeitsauszählung fand auf drei Ebenen statt (jahrgangsspezifisch; Klassenstufe 3-5):

- 1) Wort(schatz),
- 2) Phonem-/Graphem-Relationen sowie
- 3) Rechtschreibfehler.

Es sollte jeweils eruiert werden, welche Wörter in freien Texten überindividuell häufig und somit orthographisch zu beherrschen sind. Auf einer tieferen Ebene wurden die Verhältnisse von Basis- und Orthographemen betrachtet sowie deren Fehleranfälligkeiten (Band 1: Konsonanten, Band 2: Vokale (Siekmann 2023a); insgesamt 42 Einzelkapitel entsprechend den Phonem-Graphem-

Relationen: 41 plus die Graphemfolge /k+/v/). Die Fehlschreibungen wurden differenziert nach der Vertauschung 1.) *Basisgraphem für Orthographem*, 2.) *Orthographem für Basisgraphem* und 3.) *Orthographeme vertauscht*.

Zum Aufbau der Kapitel

1. Die **Kapitelreihenfolge** orientiert sich an der Häufigkeit der Einträge. Ausgehend von der seltensten Relation (in diesem Band: Phonemfolge /k+/v/, 21 Einträge) bis zur häufigsten Relation (hier: /n/, 50.281 Einträge).

2. Die Kapitel sind grundsätzlich (und dies ganz bewusst) **sprachlich und konzeptionell sehr ähnlich aufgebaut**, um eine vergleichbare und auch schnelle Orientierung zu ermöglichen (Wiederholungen wurden somit gezielt eingesetzt). Es muss nicht zwingend chronologisch gelesen werden, die interessensgeleitete Recherche zu einer Phonem-Graphem-Relation ist möglich.

3. Die jeweilige Relation wird in den (in sich geschlossenen) Kapiteln farblich hervorgehoben: **Basisgrapheme** in grün, **Orthographeme** in orange-gelb und **weitere Orthographeme**, die bei der Vermittlung der Wörter beachtet werden müssen (rötlich, nicht fett). Die Markierungen beziehen sich nur auf die Wörter der Auszählung; die Basis- und Orthographeme der Kapitelrelation werden auch im Fließtext markiert, um ein Gefühl für die Häufigkeiten anzuregen.

Am Rande: Ein Gefühl für die Häufigkeiten lässt sich über diese Form der Markierungen einfach(er) entwickeln, auch wenn Fachbegriffe wie *Phoneme/Grapheme* zu der Annahme führen könnten, dass z.B. <ph> für /f/ häufig sei.

4. Im Vergleich zu der 2011-er Auszählung (Erwachsenentexte) sind **Fremdwörter** erfasst, da sie zum Schriftwortschatz der SuS gehören (viele sind regulär nach den Prinzipien der dt. Schrift aufgebaut). **Namen** wurden dagegen ebenfalls gestrichen, da diese nicht den Konventionen der deutschen Schrift folgen müssen.

Für die Basis-/Orthographem-Statistik werden grundsätzlich die Wörter der deutschen Schriftsprache (Wortherkunft an manchen Stellen unbekannt, aber meist eindeutig; vgl. DWDS 2023) gewertet. Fremdwörter werden aufgenommen, wenn es auch Wörter aus der deutschen Schriftsprache mit dem Orthographem in der Auszählung gibt. Dies sind kleine (sicherlich auch anzweifelbare) Überlegungen der Autorin, die sich auf seltene Fälle beziehen, auf die der Fokus der Aufmerksamkeit bewusst nicht lange gelegt werden sollte.

5. Die Fehler zur **Groß- und Kleinschreibung**, die sich auf eine höhere Erwerbsebene (Satz- und später Textebene) beziehen, werden hier **nicht betrachtet** (auch nicht die Interpunktionsfehler). Dies kann zu einem späteren Zeitpunkt erfolgen. In diesen zwei Bänden geht es um die (Konsonanten-/Vokal-)Phonem-Graphem-Relationen und die jeweiligen Fehlschreibungen.

6. Für die **Zusammenschau des Wortschatzes** werden Komposita getrennt und das jeweilige **Grundmorphem** gezählt; im Fall der Phonem-Graphem-Relationen werden alle Phoneme betrachtet (aller Flexionsformen). Dies führt mitunter zu unterschiedlichen Summen und zu unterschiedlichen Platzierungen in den Einzelkapiteln.

7. Bei den jeweiligen **Top 10**-Einträgen werden zudem **Doppelplatzierungen** (Einträge mit gleich vielen Treffern) aufgeführt. Ab

Platz 11 (je weniger Einträge umso häufiger Doppelplatzierungen) wird nach alphabetischer Reihenfolge ohne Parallelzählung gewertet (an dieser Stelle macht das Alphabet als Gliederungseinheit Sinn).

8. Manche Wörter enthalten **mehrere Basisgrapheme** wie z.B. **nein** bei /n/. Die Zahlen in den Platzierungs-Kästchen beziehen sich zunächst auf das Wort, in der Addition werden mögliche Doppelungen gezählt (Doppelung sind durch einen Unterstrich gekennzeichnet). Die Summe der Top 10/20/50 enthält dann die Graphemanzahl, da sonst die Wortvorkommenshäufigkeit den Phonem-Graphem-Verhältnissen unterordnet wäre, was in der Darstellung nicht gewünscht ist.

Hierzu C.L. Naumann vor über 30 Jahren (1989: 83): „Mit Zahlen kann meistens und auch in diesem Zusammenhang nicht eine totale Festlegung, eine völlige Genauigkeit oder Sicherheit erzielt werden. Vielmehr ist es der Zweck von Zahlen, vorhandene Tendenzen, Trends, Regelmäßigkeiten zu erfassen und über das Ungefähre mehr zu sagen, als dass es ungefähr ist, kurz: Das Ungenaue genauer zu sehen.“

9. **Fugen-/Flexionsvarianten** müssen bei der Vermittlung gesondert betrachtet werden. Wichtig ist, dass ein Bewusstsein entsteht, damit Phänomene vermittelt werden können. So ist z.B. bei *vielleicht* keine Doppelkonsonanz gegeben (kein Orthographem <ll>), sondern zwei Basisgrapheme durch die Verknüpfung von **viel+leicht** (prominentes Beispiel: *Fahrrad* = **Fahr+rad**). Hierauf wird in den Kapiteln entsprechend verwiesen.

10. Zunächst wurden die korrigierten Texte gezählt. Die **Fehl-schreibungen** wurden im zweiten Schritt betrachtet und sind mit einem ***-Stern** gekennzeichnet. Die Fehlerhäufigkeiten werden in

der ersten Übersicht bereits verdeutlicht und am Kapitelende näher beleuchtet.

11. Da die didaktischen Empfehlungen hauptsächlich auf den Anfangsunterricht abzielen, werden parallel die **(An-)Lauttabellen** von rund zehn gängigen Lehrwerken/Fibeln (u.a. Cornelsen, Klett, Westermann) vergleichend betrachtet, um die Repräsentanten abzugleichen und mögliche Gegenvorschläge vor dem Hintergrund der Häufigkeiten und der Basisschreibungen anzuregen.

Grundsätzlich: Die Auszählung soll der **Orientierung** dienen. Es geht nicht um strikte Wortlisten, sondern um **das Bewusstsein um Häufigkeiten und den Aufbau der deutschen Schrift im Sinne einer logischen Vermittlungsgrundlage**. Sollte es klassenspezifisch/individuell andere Modell-/Ankerwörter (aus Basisschreibungen) geben, können und sollen diese auch genutzt werden (vermutlich finden sie sich auch unter den häufigen Wörtern).

Angesichts der Ergebnisse bleibt der **logische** Schluss:⁵

Schrifterwerb vom
„Einfachen/Häufigen“ zum „Seltenen/Schwierigen“.

Einen Überblick über die 100 häufigsten Wörter/Wortformen gibt es am Ende des Bandes (s.a. Bd. 2 zu den Vokal-Ph-G-Relationen).

⁵ Fernab aller Streitigkeiten rund um die richtige Methodik sollte das verwendete Wortmaterial sinnvoll ausgewählt sein (sinnvoll = aufbauend im Sinne der o.g. Logik; sprich: *Logik vor Methodik*).

Kurze Begriffserläuterungen

Die deutsche Schrift ist eine **lautorientierte** Alphabetschrift, d.h. die Schriftzeichen beziehen sich auf Aspekte der Lautform von Wörtern. Die lautliche und die schriftliche Ebene müssen klar unterschieden werden. Beide Ebenen haben eine feste Bezugsgröße: Beim Gesprochenen ist dies die **Standardlautung**, beim Geschriebenen ist es die **Orthographie** (vgl. Siekmann & Thomé 2018). Setzt man die schriftliche und die lautliche Ebene zueinander in Beziehung, sind es die **Phoneme**, die bedeutungsunterscheidend sind. Die deutsche Schrift orientiert sich nicht am konkret produzierten Laut (Phon), der individuell unterschiedlich ausfallen kann (z.B. durch Dialekte).

Einer bedeutungsunterscheidenden **Lauteinheit (Phonem)** steht eine bedeutungsunterscheidende **Schrifteinheit (Graphem)** gegenüber, die mitunter aus drei Einzelbuchstaben bestehen kann (<sch>, <ll>, <n>; vgl. u.a. Siekmann 2013, Thomé 2019).

Demnach gilt im Buch:

- „**Buchstaben**“ (im Sinne des Alphabets) werden in Anführungszeichen gesetzt,
- **Grapheme** stehen in < > - Klammern,
- **Phoneme** in // - Klammern.
- Ist eine lautliche Varianz im konkreten Sprachfluss (dialektaler/individueller Einfluss) gemeint, sind die sog. **Phone** durch []-Klammern gekennzeichnet.

Ein Minibeispiel:

/f/ /n/ /e:/ /b/ /a/ /l/ (6 Phoneme)
<Sch> <n> <ee> <a> <ll> (6 Grapheme, aber 10 Buchstaben)